Fantasy Sports: A Game of Skill or Chance

A Project report by:

U DINESH KUMAR*, SHAILAJA GROVER*, SHARADA SRINGESWARA*, ALOLIKA PATTANAYAK**, ABHINAV SINHA**, JIGAR PANCHAL** , JITHIN PAUL**, MOUMITA SARKER**, RAMSU SUNDARARAJAN**, TAPAN KHOPKAR**

* Indian Institute of Management Bangalore
** Cartesian Consulting Pvt. Ltd.

Executive Summary

Introduction to Online Fantasy Sports (OFS) Industry in India

Fantasy Sports is a form of skill-based online sports game where sports fans can create their own team made up of real-life players from upcoming matches. These virtual teams garner points based on the actual statistical performance of players during the real-life match and winners are determined accordingly. OFS Industry, though nascent in India, has already become immensely popular amongst sports fans with exponential growth from 20 lakh Indians playing in 2016 to over 7 Crore Indians playing in 2019. By 2020, as per industry estimates, there will be 30 crore Indian sports fans watching sports online, so we are sure that at least 33% of them (10 crore) will play fantasy sports, compared to the US where 65% of sports fans play fantasy sports. India has become the largest OFS market in the world, surpassing USA's 5 Crore Fantasy Sports user base. Dream11, an Indian unicorn is now the largest Online Fantasy Sports company in the world with over 5 Crore user base. Even the number of OFS operators has exponentially risen from 10 to 140+ in less than 3 years and expected to double by 2020.

Industry Drivers:

1. **Digital Infrastructure:** India has made exponential growth in terms of its digital infrastructure. In recent times, sports & technology have converged beautifully to provide Indian sports fans with an immersive and engaging experience, successfully bridging the offline-online sporting divide. Growing affordability of smartphones, expansion of the internet user base and plummeting data prices are fueling the growth. The internet subscribers increased from 368 million in September 2016 to 560 million in September 2018 in India, which has helped fantasy sports to flourish in the digital gaming space.

2. **User Demographics:** OFS is being consumed alike by fans across metro and non-metro cities. OFS users are mostly in the age range of 18-35 years and are tech-savvy smartphone users. Users get to play the role of the all-important manager/selector who decides the members of a team based on the statistical analysis of performance of players.

3. **Sports Leagues:** With the number of sports leagues increasing significantly, franchises across various disciplines are looking to connect with their fans in a more meaningful way. The advent of T20 leagues has further fueled more interest in the minds of sports fans with a shortened attention span and the need for instant gratification, which is why single-match freemium fantasy cricket has worked so well in India.

Indian Regulatory Landscape:

The ruling by the High Court of Punjab and Haryana and subsequent dismissal of SLP filed challenging the High Court order on Dream11's fantasy sports format as a game of skill has helped spread awareness about the legality of fantasy sports and has encouraged the participation of more users. Recently, the division bench of the Honorable Bombay High Court has also ruled that fantasy sports Dream11 format is a game of skill and does not amount to either gambling or betting.

Skills dominance - Learning effect, Consistency, Strategies and payoffs

The *dominant factor test* methodology, as used by the United States Government, has also been adopted by Indian courts to analyze the skill vs chance debate. While every game has some element of chance, where skill is dominant in determining the result of the game are considered skill based. These are games where the user's knowledge, judgment and analytical application are the more dominant than that of mere chance.

Skill dominance is established by statistically testing for *learning effect, consistency effect and impact of strategic choice on pay-offs*. **Learning effect** states that playing more regularly or practicing improves the chances of winning (improves skill) in a skill dominant game but has no effect in games dominant by chance. **Consistency effect** indicates that scores or payoffs of users in a skill dominant game will be consistent and not random (like in a game of chance). And lastly, skill-based games necessitate the use of **strategies** or a series of decisions that lead to greater payoff. In a chance dominant game, payoffs are independent of the strategies employed by the players.

This study, a first of its kind in India, employs statistical hypotheses techniques used on actual, anonymized fantasy sports data to test the above skill dominance tests. Dream11, the biggest Fantasy Sports company in India, agreed to share historical, anonymized data for the purpose of this study.

Hypothesis test as a methodology and results from Dream11 live data

The question of "Skill vs Chance" dominance was converted into an analytics & statistics problem on which hypotheses testing techniques like 2 sample t-test, Chi-square test and ANOVA were applied.

A snapshot of business hypotheses, techniques used and result

Problem construct	Skill dominance effect tested	Hypothesis	Technique used	Result
Performance of users of fantasy sports is consistent (either good or bad)	Consistency Effect	Ho: No difference in average scores of paid and free contestants (u1=u2) Ha: Average scores of paid and free contestants are different (u1 ≠ u2)	2 sample t-test	Since average scores of free and paid contestants are statistically different, we can infer that paid contestants are using some **strategy** (such as selecting high performers) and hence are **consistently** scoring more than free contestants.
Scores of the users improve as they play more number of games.	Learning Effect	Ho: There is no relationship between winning a round and the number of rounds played.	Linear regression model	Number of rounds played has a significant effect on the chances of a user to be in the top percentile of scores, hence proving a **learning** effect
Player performance has an impact on his/her selection in a team	Selection bias based on performance	Ho: Scores of strategic selection and random selection are the same Ha: Scores of the strategic and random selection are not the same	2 sample t-test	**Knowledge** and skill have a statistically significant effect on scores compared to a random selection of players to form a team.
In a game of chance, then the average percentile score will converge towards a coin-toss outcome (50-50) for all users	Non-random payoffs	Ho: No difference in average scores of users Ha: Not all the user scores are the same	ANOVA	The average scores are not equal. Specifically, there is large variance observed when one of the top 4 players is either selected or not selected as captain/vice-captain.
Selection of a player as a captain or vice-captain increases the scores of users	Impact of strategy on the payoff	Ho: Selection of players and winning (high scores) are independent of each other. Ha: Selection of players and winning (high scores) are not independent.	Chi-square test of independence	Top scorers are playing **strategically** and **choosing** players as captain or vice-captain with skill rather than randomly to achieve high scores

The tests show that users learn and fare better as they play more. They use their understanding and knowledge of the game to select players, captains, and vice-captains. With more stakes, paid users to perform better than free users indicating the usage of a strategy.

Conclusion

With various hypotheses tests, there is enough evidence to conclude that performance in fantasy sports is skill dominant. There is strong analysis that provides evidence of learning, consistency and strategy-based usage all of which affect payoffs. There is little to no evidence of randomness or chance. This case has already been published in Harvard Business Publication and is available across more than 410 universities across 75 countries. The teaching note and case study is also being used by Prof U Dinesh Kumar (protagonist author) as part of his student curriculum in IIM Bangalore.

भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT BANGALORE

CARTESIAN
The Precision Practice

Project Report

INTRODUCTION

Fantasy sports is the fastest growing online gaming industry worth several billion dollars. An important debate associated with fantasy sports across the world is whether it is a game of skill or chance. In India, online gaming laws are not very well defined. While the game of skill is permitted by Indian laws, a game of chance involving money is strictly prohibited. Indian courts have recognized that no game is a game of pure skill or game of pure chance.[1] When there is a chance involved, the Indian courts decide based on the dominant factor test[2] that requires determining whether the skill is dominant, or chance is the dominant factor for a given instance. Can this debate around whether the dominant factor is a skill or chance be settled using data? With this objective, Dream11 which is one of the largest fantasy sports companies in India was approached to understand the game as well as to collect data to test some of the hypotheses which can help differentiate if in fantasy sports the dominant factor is skill or chance. Dream11 has created fantasy games for sports such as cricket, football, hockey, kabaddi, and basketball.

The main concern for lawmakers across the world was whether fantasy sports was a game of skill or game of chance. Edelman in 2016 claimed that fantasy sports were a nerds' game[3]. Similarly, Jennifer Chu[4] claimed that fantasy sports involve real skill. Authors such as Edelman and Jennifer Chu believed that the players of fantasy sports would need a better understanding of the sports, how the players perform under different playing conditions, injury to players, and the impact of weather and other playing conditions. For example, till June 2019 Rafael Nadal won 12 of his 18 grand slam titles in the French Open, whereas Roger Federer won just one French Open out of his 20 grand slam titles. This is an example to show that fantasy sports enthusiasts should be aware of the strengths and weaknesses of players before selecting them. Through our study we want to see if data could be used to check whether fantasy sports was a game of chance or skill, especially whether skill is a dominant factor in winning fantasy sports competition as claimed by Edelman.

---

1 Source: Supreme Court of India, Dr K R Lakshmanan Vs State of Tamil Nadu and ANR on January 12, 1996, available at https://indiankanoon.org/doc/1248365/

2 G Gokhale *et al.* (2018), "The laws related to Fantasy Sports Games in India", *The Sports Law & Policy Centre,*

3 M Edelman, "A Sure Bet? The Legal Status of Daily Fantasy Sports', *Pace Intellectual Property, Sports Entertainment Law Forum",* **6**(1), 1-21, 2016.

4 Jennifer Chu, "There is real skill in fantasy sports", *Science Daily,* November 7, 2018. Accessible at https://www.sciencedaily.com/releases/2018/11/181107093816.htm.

ABOUT THE FANTASY SPORTS INDUSTRY

Fantasy sports is was a competition game among fans that involves building an imaginary team comprising players for upcoming real match game. Based on the combined l performance of the players in a match, the fans are awarded points and they can win a prize[5]. History of fantasy sports can be traced to the 1960s when Bill Gamson, former Professor of Psychology at Harvard and Michigan University constructed a game called "The Baseball Seminar"[6]. Daniel Okrent, student at the Michigan University learned about "Baseball Seminar" and created his own version of the fantasy sports called "rotisserie baseball,[7]" whose idea had been crystallized in a Manhattan restaurant named La Rotisserie Francaise[8].

In 2018, the fantasy sports market worth was USD 13.9 billion and expected to grow at a CAGR of 13.7% reaching USD 26.4 billion in 2024, according to a study carried out by Market Study Report LLC[9]. A study conducted by Fantasy Sports and Trade Association, claimed that the fantasy sports users grew at 25% annually[10]. With such high anticipated growth, channels such as ESPN broadcast hour-long shows devoted to fantasy sports and players prior to major sports seasons.

In the past 5-6 years, fantasy sports has become one of the most popular activities for sports lovers in India. The major reason for this surge in fantasy sports users could be the rise of league-based sports, and the growth of online viewership combined with increased smartphone penetration. Technological advancement has also contributed to the growth in popularity of

---

5 Source: David O Klein et al., "Fantasy Sports – Rapidly Developing Legal Framework", *Law 360,* September 20, 2017, accessible at https://www.law360.com/articles/704275.

6 Justin Fielkow, "From Fantasy Sports to Reality: The evolution and legality of fantasy sports", *The Sports Esquire,* May 18, 2015 available at https://charlesfranklinlaw.com/wp-content/uploads/2016/12/2015.05.18-From-Fantasy-to-Reality_-The-Evolution-and-Legality-of-Fantasy-Sports-Fielkow-The-Sports-Esquires-Revised.pdf

7 Ibid

8 Source: History and Evolution of Fantasy Sports available at http://futureoffantasy.com/the-history-and-evolution-of-fantasy-sports

9 Source: https://www.marketwatch.com/press-release/at-137-cagr-fantasy-sports-market-size-is-anticipated-to-reach-us-26400-million-by-2025-2019-03-29

10 Source: D Heitner, "The Hyper Growth of Daily Fantasy Sports is Going to Change Our Culture and Our Laws", *Forbes,* September 16, 2015. Available at https://www.forbes.com/sites/darrenheitner/2015/09/16/the-hyper-growth-of-daily-fantasy-sports-is-going-to-change-our-culture-and-our-laws/#3331b7615aca

fantasy sports in India. While cricket has dominated the sports market in India, the new generation is exploring and getting interested in other sports such as Kabaddi, football, basketball and hockey.

The Indian fantasy sports market has shown a double-digit growth. A report by AC Nielsen, commissioned by the Indian Federation of Sports Gaming (IFSG) across 12 cities in India claimed that 67% of the survey respondents were aware of fantasy sports and the retention rate of users was 89%[11]. The same report claimed an increase in number of users by 200% in 2017. While the US and Canada had 59 million users in 2018, the Indian market saw an increase in users from 2 million users in 2016 to 20 million in 2018, more than the number of users in the UK[12]. An interesting fact is that fantasy sports was witnessing 40% growth in female user participation during the International Cricket Council (ICC) World Cup[13]. In 2019, India had more than 80 million fantasy sports users across 130+ different platforms, which was expected to grow to 100 million users by 2020[14].

Fantasy sports and the gaming industry was set to witness a massive growth in India, growing from Rs. 43.8 billion to reach Rs 118.8 billion by FY23 at CAGR 22.1%, according to a report by Indian Federation of Sports Gaming (IFSG), India's first and only self-regulatory industry body for the sports gaming sector and KPMG India Private Limited (KPMG). The report titled 'The Evolving Landscape of Sports Gaming in India' provides an overview of the online gaming industry with a focus on fantasy sports and eSports[15].

FANTASY SPORTS COMPANIES IN INDIA

There were around 150+ online fantasy sports gaming companies in India such as Dream11, MyTeam11, HalaPlay, BalleBaazi, Fanmojo, and 11Wickets. Dream11 is the biggest fantasy sports company in India. Dream11 was co-founded by Harsh Jain and Bhavit Seth in 2008; and in 2012, they introduced the freemium format of fantasy sports to the Indian cricket fans. In 2014, the company reported 1 million registered users, which grew to 2 million in 2016 and to 60 million in 2018[16] [17].

---

11 Source: "Scoring Big with Sports Gaming", A report on emergence, consumption, patterns, meteoric growth and future of fantasy sports in India", available at https://www.ifsg.in/wp-content/uploads/2018/03/IFSG-Report_33-low-res.pdf

12 Ibid

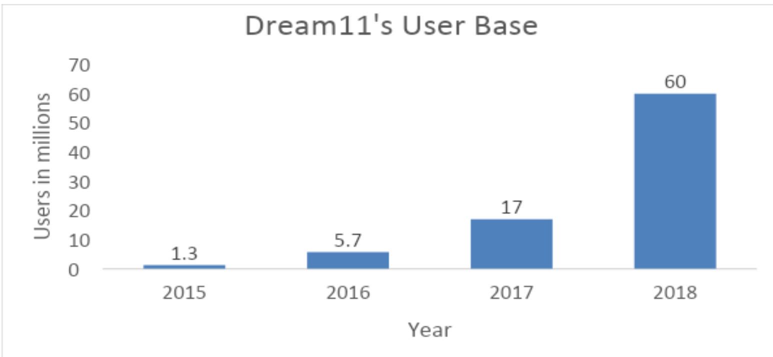
13 Ibid

14 Source: Sporting Revolution in India available at https://www.thehindu.com/brandhub/a-sporting-revolution-in-india-tracking-the-growth-of-fantasy-sports/article24026469.ece

15 Source: The evolving landscape of Sports Gaming in India available at https://www.ifsg.in/publications/

16 Source : https://www.outlookindia.com/newsscroll/fih-announces-partnership-with-dream11/1430388

17 Source : https://scroll.in/article/919582/big-game-how-a-fantasy-sports-startup-is-making-money-from-indias-ipl-fever



In 2019, it was reported that Dream11 held 90% of the domestic fantasy sports market[18]. In 2018, Dream11 signed MS Dhoni, India's ex-captain as their new brand ambassador and subsequently launched a media campaign under the banner "Dimaag se Dhoni" during Indian Premier League 2018[19].

HalaPlay was founded by the students of Birla Institute of Technology (BITS) Pilani in 2017. In 2019, HalaPlay reported 4 million users and 10 times growth over the last 12 months[20]. MyTeam11, a Jaipur based startup was founded in 2016 by two engineers and had around 10 million+ users.

ONLINE GAMBLING LAWS ACROSS THE WORLD

In USA, fantasy sports were exempted from federal law concerning the Unlawful Internet Gambling Enforcement Act (IGEA)[21]. It was left to a state-level decision; each state applied various standards to determine if fantasy sports were a game of skill or chance. The decision between skill and chance was to be taken based on whether the skill-based element was dominant over chance in determining the outcome of the game[22].

[18] Source: The Rise of Fantasy Sports in India and Startups to Watch for in the Space, available at https://yourstory.com/2019/04/fantasy-sports-startup-dream11-mpl-ipl

[19] Source : https://en.wikipedia.org/wiki/Dream11

[20] https://yourstory.com/2019/04/fantasy-sports-startup-dream11-mpl-ipl

[21] Jonathan Griffin, "The Legality of Fantasy Sports", *National Conference of State Legislatures",* **23**(33)**,** September 2015.

[22] Source : https://en.wikipedia.org/wiki/Gambling_in_the_United_States

Most European countries recognized fantasy sports as a legal business; it was legal and highly regulated in around 30 countries in Europe. In Sweden, Croatia and Norway, though fantasy sports was legal, it was not very regulated. However, in countries such as Switzerland, Ukraine, Macedonia and Belarus, online gaming was considered illegal[23].

Sports & Entertainment and Gambling & Betting have been a state subject in India and respective states can formulate laws for various gaming activities. Most of the laws governing the subject of gambling in India in 2019 were old. The said laws used a yardstick of 'Skill vs. Chance' and predominance of skill over chance to identify if a given game falls within the ambit of gambling. Advent of online games coupled with lack of awareness about them, created confusion as to how this yardstick of skill versus chance or game of mere skill could be applied. The two main enactments dealing with gaming in India were the Pre-Independence Public Gambling Act, 1867 ("PGA") and the Prize Competitions Act, 1955 ("PCA"). In general, gambling laws of most states prohibited gambling, but the laws created a carve out for games involving 'mere skill' from the applicability of provisions of gambling Act.' Historically, Indian laws have differentiated between the games of skill and games of chance. While the game of skill was permitted by Indian laws, game of chance was strictly prohibited.

SKILL BASED VS. CHANCE BASED

Games of skill are the games which involve a person's skill, knowledge and judgment and they rule out chance aspect from the game. On the other hand, games of chance include games that are determined by mere luck, completely uncertain and the players cannot apply their skill to estimate the result[24].

While deciding whether a game is a game of chance or skill, the Indian courts while interpreting the term "mere skill" have adopted the methodology used by the US government, that is, the "dominant factor test". It says that while classifying a game, it becomes important to check whether chance or skill is the dominant factor in determining the result of the game. Dominant factor test was applied to Rummy and Horse Racing. The Supreme Court of India classified Rummy as a game of skill, as the fall of cards need to be memorized and considerable skill is required while deciding whether to hold or discard a card[25]. On similar lines, horse racing requires objective assessment of fitness and skill of the horse and the jockey.

---

[23] Source : https://blog.vinfotech.com/fantasy-sports/european-laws-on-fantasy-sports

[24] Source : https://www.ifsg.in/wp-content/uploads/2018/03/FantasySportsPublication.pdf

[25] Source: K R Lakshmanan Vs State of Tamil Nadu, available at https://www.sci.gov.in/jonew/judis/16203.pdf

In 2018, the Supreme Court of India concurred with earlier verdicts by Punjab and Haryana High Courts that Fantasy sports involves substantial amount of skill.[26] The High Court of Punjab and Haryana ruled that fantasy sports was predominantly a skill-based game. It is pertinent to note here that it is difficult to conceive of any game of skill as a game of pure skill and no chance. Some element of chance is always present in any game of skill. For instance, almost all forms of popular sports involving very high degrees of skill, such as chess, cricket, football, hockey, etc. also involve some element of chance. It becomes important to obtain evidence from data by developing statistical models to test and classify fantasy games as a game of skill or game of chance. If a fantasy sports is chance based, then every user should have an equal probability of winning, whereas if it is skill based, then one should see consistent performance among the users. The only way to check this is to use the data from a fantasy sports company.
To undertake this study Dream11 was approached with a request to share anonymized data for testing whether fantasy sports involving cricket was skill dominant or chance dominant. Dream11 agreed to share their data and rules related to fantasy sports based on a cricket series, which are described in sections below.

DREAM11: LET'S PLAY (Rules and constraints)

The following steps were to be adopted to play a game in Dream11.

- Select A Match: Select any of the upcoming matches from any of the current or upcoming cricket series.
- Create Your Team: Create a team with all available players for a match with the following constraints.
  - Each player comes at a price, a budget of 100 credits is available to create a team of 11 players.
  - User can create multiple teams and choose to join a contest with any of the teams created.
  - After creating your team on Dream11 platform, choose a Captain and Vice-Captain of your choice for the team.
    - Captain – Gets 2x points scored by him in the actual game
    - Vice-Captain – Gets 1.5x points scored by him in the actual game
  - There will be multiple contests for a match, User needs to select a contest and assign a team.

[26] Source: G Gokhale and Rishabh Sharma, "The skill element in the fantasy sports games", *The Sports Law & Policy Centre", available at*
http://www.nishithdesai.com/fileadmin/user_upload/pdfs/NDA%20In%20The%20Media/News%20Articles/180406_A_Legality_of_Fantasy_Sports_India.pdf

- You can make as many changes to your teams created on Dream11 platform as you wish until the deadline of that match. You can also change your Captain or Vice-Captain before the deadline of the match.
- Contests would be open for submitting a team till the deadline of that match. There can be no changes to submitted team or possibility of adding new team after the start of a match.
- Team submission can be done till slots are available in a contest.
- Create your Dream11 team by picking 11 players as per the following combination C2, ns (C1, C2, C3...) within a budget of 100 credits. Maximum number of players from a team cannot exceed 7. In other words, your team should necessarily consist of players from both the real game participating team.

Player type	C1	C2	C3	C4	C5	C6	C7
Wicket Keeper	1	1	1	1	1	1	1
Batsman	5	5	4	4	4	3	3
All Rounder	2	1	1	2	3	2	3
Bowler	3	4	5	4	3	5	4
Total	11	11	11	11	11	11	11

Dream11 :DATA HIERARCHY

The data available for analysis was provided in various tables. The overall hierarchy of data is as follows.

- Various rounds/match were played. For example, one IPL match would be one round.
- There were players who were available to be picked up for a round or match. (This number is more than the number of players who would play in that match, so it was possible that a player selected by a user in his team may not actually play.)
- For every round, multiple contests were opened. The contests were of different categories, from free to paid, and various types of playing and winning options (public, private, special).
- User selected a team for a round and for a contest.
- There was a player round performance table which indicated how the player performed in the specific round.
- Teams selected by users were scored based on the selected player's performance in a contest and those team level scores were provided in the contest user's table.

Data Description:

Brief description of the data provided by Dream11

- Data tables
  - CONTEST MASTER – Provides details of a contest such as if the contest is free or paid, public or special or private. It also has details on the size of the contest.
  - CONTEST USER RELATION – Provides details points scored by a user in a contest for the team which has been fielded.
  - PLAYER ROUND PERFORMANCE – It has details on performance of a player in a round.
  - ROUND MATCH – Association between roundid and matchid which denotes number of matches played for a specific round or game. The scores are given for a match, hence to get the scores for a user team, we need to map roundid with matchid using this table.
  - USER ROUND PLAYER – This table has all the players selected by user or the team which is formed by the user for a round.
  - USER ROUND TEAM – This table contains the date on which the user formed the team, this table is mainly used for joining all other tables since it contains all the required information about a round→user→team.
  - AGGREGATE_DATA – This table has aggregate data which provides details of user's performance across multiple rounds.

- Attributes description
  - roundid – Unique identifier for round played
  - matchid – Unique identifier for matches played in a round
  - contestid – Unique identifier for a contest. A round can have multiple contests.
  - Contesttype – Public/Special/Private
  - Contestcategory – Free/Paid
  - teamid – Unique identifier for a team. User can create multiple teams for a round.
  - currentpoints – Points scored by the user for the selected team in a contest for a round
  - userid – Unique identifier of a user

भारतीय प्रबंध संस्थान बेंगलूर
INDIAN INSTITUTE OF MANAGEMENT
BANGALORE

CARTESIAN
The Precision Practice

Details of data in all the tables provided by Dream11:

2 Rounds 5365 & 6349 have been considered. These 2 rounds are having high numbers of users playing the game.

- 5365 - Is a T20 Premier League match played on 29/05/16?
- 6349 - Is a ODI IND vs NZ match played on 29/10/16?

File: USER_ROUND_PLAYER.CSV

This table has information on team formed and fielded (11 players along with captain and vice-captain) by the user for a round/match

- It has 531425 rows with 17 columns.
- 2 unique round ids
- 3006757 unique users
- Users having as many as 4 teams per round

File: CONTEST_MASTER.CSV

This table has information on the contest which is open for a round/match. It also has details on type and size of each contest.

- It has 74220 rows and 5 columns.
- 2 unique round ids
- 74220 unique contests
    - 5365 has 42247 unique contests
    - 6349 has 31973 unique contests
- 4 unique contest types (public, private, special, grand)
- 2 contest categories (paid and free)

File: CONTEST_USER_RELATION.CSV

This table has information on points scored by a user for a team formed, in a round/match and contest combination.

- It has 962602 rows and 6 columns.
- 2 unique round ids
- 72652 unique contest
- 239670 unique user ids
- 4 unique contest types (public, private, special, grand)
    - special - 699330 records
    - public - 160271 records
    - grand - 84051 records
    - private - 18950 records
- Max number of points scored by a user is 447.25, minimum being 0.
- Max number of teams formed by a user is 4.

File: PLAYER_ROUND_PERFORMANCE.CSV

This table has information on points scored by a player in a round/match. It has detailed break-up of how many points player scored for bowling, batting and fielding.

- It has 44 rows and 43 columns.
- 2 unique match ids
- 43 unique players

- Highest point scored by a player is 72, minimum being -1.

File: round_match.csv

This table has mapping between matchid and roundid.

- It has 2 rows and 2 columns.
- 2 unique match ids
- 2 unique round ids

File: USER_ROUND_TEAM.CSV

This table has information on date of match, user and round/match.

- It has 285630 rows and 4 columns.
- 2 unique round ids

File: AGGREGATE_DATA.CSV

This table has aggregated information of user's performance across multiple rounds.

- It has 385235 rows and 17 columns.
- 385235 unique user ids

METHODOLOGY

The main methodology used to test if Fantasy Sports is a game of skill or chance using data provided by Dream11 was to create a couple of hypothesis which tests if Fantasy sports users are applying any skill in choosing a team for a game and which gets reflected in their scores. Is there any learning affect which is reflected in the scores of players playing regularly? These assumptions are tested by creating various hypothesis tests and getting appropriate data sets from the entire data provided by Dream11. Testing those hypotheses on these data sets created by slicing and dicing the data on various parameters using different hypothesis tests help us reach conclusions which are data driven.

Few possible hypotheses which helps us test skill vs chance are:

1. Users playing free contests are scoring lower than users playing paid contests. This can prove that when users play paid contests, they play more cautiously and strategically, and do not select teams at random.
2. Scores of randomly selected players can be tested against scores of the teams based on a specific strategy such as selecting players who have performed well in the recent matches.
3. Is the selection of captains and vice captains of the team random (equal probability)?
4. Selection of players and winning or getting high scores are dependent on each other.
5. As the user plays more games, his chance of winning increases (learning effect).

Various hypothesis tests and inferences

First of all this question of "Skill vs Chance" can be converted into a data analytics problem where techniques like hypothesis testing helps in solving the dilemma. The output of fantasy sports is in the form of scores of individual users based on their team selection, which eventually gets converted into the money they earn. Scores will be our key variable (outcome variable), which will drive whether the scores are random or are an output of a more strategic team selection. The following are the possible approaches for testing skill vs. chance.

a. Consistency in performance of users of fantasy sports, that is, if it is skill dominant, the scores will be consistent.
b. Selection of players with best performance in the recent matches and the scores of users.
c. Use of strategy such as selecting players based on past performance, as vice-captain and captain.
d. Improvement in scores as the number of games played by users increases.
e. Average scores whenever a player is chosen or not (dependency of scores on selection of a specific player).

Few sets of hypotheses as listed below can help us test the concept of "Skill Vs Chance". Below table explain how these hypotheses are linked to our problem and how the data provided can help us test these hypothesis

Table 1 - Hypotheses for checking whether fantasy sports is skill based or chance based

Hypothesis	Link to skill vs. chance	Constructing the hypothesis test using the data
Performance of users of fantasy sports is consistent (either good or bad)	*In a game of chance, performance will be inconsistent, whereas in a game of skill, the performance will be consistent.*	Users in the top quartile of multiple rounds can be chosen and tested on the percentage of times they have been in the top quartile.
Scores of the users improves as they play more number of games.	*Continuous learning is a very important aspect of skill. We can check whether there is evidence that with more number of games played, the score of a user improves.*	This hypothesis can be tested using two approaches. (1) Build a regression model and check if scores of a user improve with number of games played. (2) Conduct a chi-square test of independence for scores and number of rounds by creating appropriate bins.

Player performance has an impact on his selection in a team.	*This can be used to prove that the selection of player is not random.*	Data on player performance can be correlated with number of users selecting the player.
Average percentile score of users is not 0.5.	This can be used to prove that the fantasy sports is not a game of chance.	If fantasy sports is a game of chance, then the average percentile score will converge towards 0.5 for all users.
Selection of a player as a captain or vice-captain increases the scores of users.	Skill is required to choose the right player as captain or vice-captain	ANOVA can be used to check average scores of users and their selection of captain/vice-captain.

These are few of the claims, there can be many more. We will test a few using the data provided.

1. First Hypothesis : As we know Dream11 platform has both free and paid users, the users who play games for free with no return and users who pay a fee and obtain returns at the end of the game based on their relative performance. If we can statistically test that scores of paid users are significantly higher than the scores of free users, then we can say users who play for money play a more strategic game and score more. To test this hypothesis, we will compare the average scores of paid and free contest users and check if there is a statistically significant difference between their average scores.

We will create two samples of data, one of the scores of paid contest users and another of scores of free contest users and conduct a 2-sample t-test with unequal variance to test the hypothesis.
Ho: No difference in average scores of paid and free contestants ($\mu_1 = \mu_2$)

Ha: Average scores of paid and free contestants are different ($\mu_1 \neq \mu_2$)

where μ_1 is the average score of players who play free contests; and μ_2 is the average score of players who play paid contests. The number of users playing free and paid contests is shown in **Figure 1**. The distribution of scores of paid and free contests are shown in **Figure 2**.The hypothesis result output is shown in **Table 2**. The p-value is very small; so we reject the null hypothesis, which means the average scores of paid users playing paid contests is more than the average scores of players playing free contests.

Figure 1

Number of users of free and paid contests for rounds 5365 and 6349

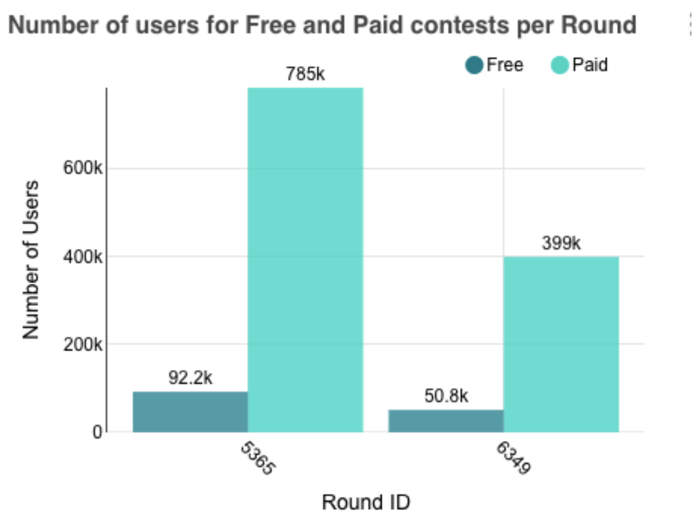


Figure 2

Distribution of scores of free (red line) and paid contests (blue line)

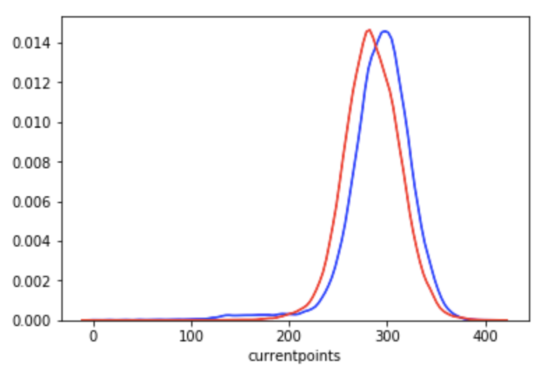


Table 2

Hypothesis Test Result

t-stat value = 18.35175195627833
p value = 3/2088574221489635e-75
Mean of Paid contest type = 236.87613622680416
Mean of Free contest type = 235.16216597368555
Standard Deviation for Paid contest type = 133.3065137899964
Standard Deviation for Free contest type = 130.34312313429582

Key inference out of this test is - Since average scores of free and paid contestants are statistically different, we can infer that paid contestants are using some strategy (such as selecting high performers) as money is involved and hence are scoring more than free contestants.

2. Second Hypothesis is if the scores of users who use some strategy to select players such as recent performance of players is higher than users who select players randomly. This hypothesis is to test whether scores obtained by teams selected based on strategy and knowledge of players' past performance are higher than scores obtained by teams with random selection of players.

In order to test this, we need to create two samples as described below:

- One sample is based on random selection of players from the available list of players for the game and applying all the constraints of fantasy sports while creating the teams.
- Second sample is created by selecting the team based on the past performance of the available players. Data of performance of two teams (in this instance, Gujarat Lions and Sunrisers Hyderabad IPL teams were considered). Data on performance of all the players of these two teams in the IPL matches was collected from the portal CricInfo. Based on the performance, top players were selected for batsman, bowler, allrounder, wicket keeper, captain and vice-captain categories.

For the second sample, we formed teams based on the following strategies:

- Team composition: 1 All-rounder, 5 Batsmen, 4 Bowlers, 1 wicket keeper, captain and vice-captain
- Team composition: 3 All-rounder, 4 Batsmen, 3 Bowlers, 1 wicket keeper, captain and vice-captain

- Top players based on their performance in the previous matches were randomly selected and teams were formed using the selected strategy. In this sample, the teams were formed with all the constraints, but players were selected only from top performers of the previous matches.
- During the next match of these two teams (Gujarat Lions and Sunrisers Hyderabad) based on the performance of these players in the subsequent match, scores were given to all the teams in the two samples (teams with randomly selected players and teams with players selected based on knowledge).
- 2-sample t-test was conducted on these two samples of scores obtained by these teams.

Ho: Scores of strategic selection and random selection are same.
Ha: Scores of strategic and random selection are not same.
Number of records in strategy users selection set = 36
Number of records in random selection set = 36
Results are as follows:
t-stat value = 11.331975795254104
p-value = 2.056343133916253e-17
Mean of strategy-based selection dataset = 416.6666666666667
Mean of random selection dataset = 240.95833333333334
Standard deviation for strategy-based selection dataset = 68.490368341
Standard deviation for random selection dataset = 61.02314849938323

The p-value of the hypothesis tests is very low; hence, we reject the null hypothesis and conclude that the average scores of selecting players based on the recent performance is more than the random selection of players.

Key Inference out of this test is - that knowledge and skill have statistically significant effect on scores compared to random selection of players to form a team.

3. Third hypothesis - If fantasy-sports is a game of skill, then player performance plays a major role in the player getting selected to a team as well as selection of captain or vice-captain. If by using the data, we can test if selection of players in a team and getting high scores are dependent on each other we will be able to prove that users are selecting their team players based on their skill of knowing and understanding past player performance.

In this hypothesis, we are assuming that the right selection of players increases the chance of higher scores for a user and hence selection of players and scores are not independent of each other. Similarly, users with high scores have chosen high performing players as captain or vice-captain.

Ho: Selection of players and winning (high scores) are independent of each other.
Ha: Selection of players and winning (high scores) are not independent.

We can perform a Chi-square test of independence to test whether selection of players and scores are independent. In this instance, we select one player and create a contingency table listing the number of times the player was selected in a team by users and number of times the scores of these users were in the top quartile. Similarly, number of times a player was not selected by users and the score was in the top quartile was listed.

Player ID 635: (We have chosen a specific high-performance player and tested the hypothesis for this player being selected as a team member and also as a captain and its impact on the scores of the users who selected him). The contingency table is provided in tables 3-4 below.

Table 3

Selection of player ID 635 and scores

Round id	6349			
player Id	635			
	Selected	Not Selected	**Total**	
TopQ	51558	1464	**53022**	
Others	152036	14821	**166857**	
Total	**203594**	**16285**	**219879**	

Segment	**Class**	**Oij**	**Eij**	**(Oij - Eij)^2/Eij**
TopQ	Selected	51558	49095.0071	123.5631552
Others	Selected	152036	154498.993	39.26455358
TopQ	NotSelected	1464	3926.99289	1544.778448
Others	NotSelected	14821	12358.0071	490.8828689
		Chi-square Statisctic =		**2198.489026**

Table 4

Player ID 635 selected as captain

Round id	5365			
Captain id	635			
	Selected	Not Selected	**Total**	
TopQ	33133	25203	**58336**	
Others	86660	100829	**187489**	
Total	**119793**	**126032**	**245825**	
Segment	**Class**	**Oij**	**Eij**	**Oij - Eij)^2/Ei**
TopQ	Selected	33133	28427.7207	778.80507
Others	Selected	86660	91365.2793	242.320203
TopQ	NotSelected	25203	29908.2793	740.251648
Others	NotSelected	100829	96123.7207	230.324553
		Chi-square Statisctic =		**1991.70147**

The chi-square statistic values from above tables are very high. Hence, we reject the null hypothesis that the scores of fantasy sports users and selection of specific players in the team are independent of each other.

Similar chi-square test of independence was carried out for several players and chi-square statistic values are shown in table 5 below. In all these instances, the null hypothesis is rejected.

Table 5

Chi-square test of independence for several players

Not selected-low quartile	not selected top quartile	playerid	roundid	selected low quartile	selected top quartile	Total_notSelected	Total_Selected	Total_TopQ	Total low quartile	All_total	E_topQ_sel	E_LowQ_sel	E_topQ_notsel	E_LowQ_notsel	Chi-sqr_val
14821	1464	635	6349	152036	51558	16285	203594	53022	166857	219879	49095.0071	154498.993	3926.99289	12358.0071	2198.48903
16198	1711	635	5365	171291	56625	17909	227916	58336	187489	245825	54086.0684	173829.932	4249.93155	13659.0684	2144.97156
11754	5603	635	5360	148535	46173	17357	194708	51776	160289	212065	47538.2614	147169.739	4237.73858	13119.2614	633.79375
11426	2559	1	6348	118279	41765	13985	160044	44324	129705	174029	40762.1158	119281.884	3561.88417	10423.1158	411.973132
7106	1875	342	5118	139743	41847	8981	181590	43722	146849	190571	41661.5224	139928.478	2060.47763	6920.52237	22.7387122
4435	617	635	5109	116978	38455	5052	155433	39072	121413	160485	37842.0299	117590.97	1229.97005	3822.02995	416.912053
19172	14	119	5364	138722	52022	19186	190744	52036	157894	209930	47280.3067	143463.693	4755.69331	14430.3067	6918.08
12023	1822	119	5361	143897	46434	13845	190331	48256	155920	204176	44983.8019	145347.198	3272.1981	10572.8019	902.844807
4446	780	6	5111	108926	35612	5226	144538	36392	113372	149764	35122.1048	109415.895	1269.89525	3956.10475	258.681618
8267	149	119	5110	111767	36749	8416	148516	36898	120034	156932	34919.2221	113596.778	1978.77787	6437.22213	2337.46541

Key Inference out of this test is - This hypothesis test establishes that player selection has a huge influence on achieving high scores. Top scorers are playing strategically and choosing players with skill rather than randomly to achieve high scores.

4. Fourth Hypothesis – We can also test the reverse hypothesis than what we tested in the previous test. If choosing a poor performing player as a captain or vice captain has any affect on the scores of the users. We can test the above hypothesis by checking that users who selected poorly performing players as captain or vice captain and their scores are dependent on each other.

In this hypothesis, we are assuming that the poor selection of players reduces the chance of higher scores for a user and hence selection of players and scores are not independent of each other. Similarly, users with low scores have chosen poor performing players as captain or vice-captain.

Ho: Selection of players and loosing (poor scores) are independent of each other.
Ha: Selection of players and loosing (poor scores) are not independent.

We can perform a Chi-square test of independence to test whether selection of players and scores are independent. In this instance, we select one player who performed poorly in previous games and create a contingency table listing the number of times the player was selected in a team by users and number of times the scores of these users were in the top quartile. Similarly, number of times a player was not selected by users and the score was in the top quartile was listed. For this test we chose poor performing players from past data and tested hypothesis on sample data where those players are selected in the team and selected as captain of the team.

Player ID 111: (We have chosen a specific poor-performance player and tested the hypothesis for this player being selected as a team member and its impact on the scores of the users who selected him). The contingency table is provided in tables 6 below.

Table 6 :

Round id	6349			
player Id	111			
	Selected	Not Selected	**Total**	
TopQ	101	52921	**53022**	
Others	2708	164149	**166857**	
Total	**2809**	**217070**	**219879**	
Segment	**Class**	**Oij**	**Eij**	**(Oij - Eij)^2/Eij**
TopQ	Selected	101	677.367088	490.4268693
Others	Selected	2708	2131.63291	155.8425086
TopQ	NotSelected	52921	52344.6329	6.346381701
Others	NotSelected	164149	164725.367	2.01668405
			Chi-square Statistic =	**654.6324436**
			DF =	**1**

Player ID 731: (We have chosen a specific poor-performance player and tested the hypothesis for this player being selected as a captain and its impact on the scores of the users who selected him). The contingency table is provided in tables 7 below.

Round id	5365			
Captain id	731			
	Selected	Not Selected	**Total**	
TopQ	0	58336	**58336**	
Others	65	187424	**187489**	
Total	**65**	**245760**	**245825**	

Segment	**Class**	**Oij**	**Eij**	**(Oij - Eij)^2/Eij**
TopQ	Selected	0	15.4249568	15.4249568
Others	Selected	65	49.5750432	4.79937638
TopQ	NotSelected	58336	58320.575	0.00407968
Others	NotSelected	187424	187439.425	0.00126937
			Chi-square Statistic =	**20.2296822**
			DF =	**1**

The chi-square statistic values from above tables are very high. Hence, we reject the null hypothesis and can confirm that scores of fantasy sports users and selection of specific players in the team are dependent of each other. In both cases poor performing players were selected and we saw that users who selected those players in their team and their chances of being in the top quartile is very low, especially in case of captain selection. First of all, very few people have selected poor performing player as captain and whosoever has selected has never been in top quartile.

Key Inference out of this test is - This hypothesis test establishes that player selection has a huge influence on achieving high scores. Top scorers are playing strategically and choosing players with skill rather than randomly to achieve high scores. Poorly selected teams can never bring a user in top quartile of scores.

5. Fifth Hypothesis – In order to test that selecting top performing players in your team can increase the chances of getting high scores we need to check if the top scorers of Fantasy sports are choosing top performing players and that will prove that skill is needed to get to the top scorer spot.

This hypothesis aims to test if the average scores of users selecting top players and not selecting top players are equal. In a game of skill, this hypothesis will be used to establish the fact that high score is dependent on the team which is selected by the users and not random.

For one top player, we will divide the dataset into two samples; one with the scores of users who selected the player in the specific rounds where he played and the other sample with the scores of the users who did not select the player in the specific round and we perform a 2-sample t-test. **Figure 3** shows average scores of users who have selected player ID 635 as Captain/Vice-

Captain. **Figure 4** shows the number of users who have selected player ID 635 as Captain/Vice-Captain.

Figure 3
Average scores of users when player ID 635 was either selected (Y) or not selected (N) as captain/vice-captain for rounds 6349 and 5365

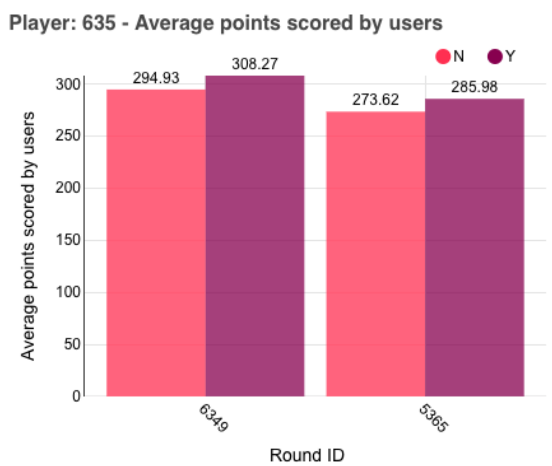
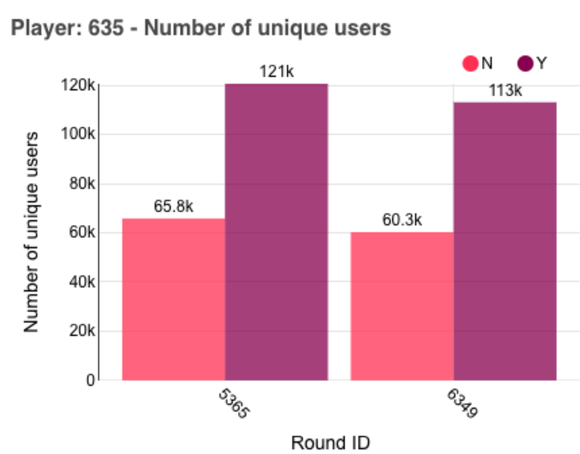


Figure 4
Number of users who have selected (Y) or not selected (N) player ID 635 as captain/vice-captain for rounds 6349 and 5365



Ho: Average scores of users who selected the top player and not selected are equal.
Ha: There is a difference between the scores of the users who selected the top players and who did not select them.

To test this hypothesis, the following process is used

- Top 20 high scoring players are selected
- For each round
  - Matches or Round in which the players are selected
  - For each round in which the top players played
    - Select users who participated
    - Two samples created with the scores of users who selected the players and users who did not select the players
    - 2-sample t-test conducted for the 2 samples created from multiple rounds

The results of hypothesis test are shown in table -7 below.
Table 7

Results of hypothesis test

playerid	roundid	Avg_Score_when_not_selected	standard_deviation_notSelected	Avg_Score_when_selected	standard_deviation_Selected	t_stat	p_val
1091	3892	287.60	124.32	554.02	60.94	4.67	0.01
432	3896	813.73	100.54	901.99	119.74	16.27	0.00
1343	6227	231.71	60.00	456.92	99.50	9.45	0.00
1236	6532	461.64	100.80	596.24	154.67	31.62	0.00
309	2499	345.52	96.66	560.48	117.39	17.59	0.00
309	2452	261.75	74.39	473.84	93.93	108.83	0.00
309	2500	284.24	78.91	507.46	98.45	22.18	0.00
342	5060	311.81	93.44	498.15	95.05	13.27	0.00
342	5059	345.55	108.37	543.65	111.91	12.76	0.00
1443	3662	271.18	120.93	339.75	34.29	1.80	0.13
1021	6354	545.22	80.43	705.92	96.97	129.18	0.00
1885	6408	499.53	143.36	693.16	137.22	30.18	0.00
259	4725	263.03	60.79	506.84	96.97	8.10	0.00
1161	4213	446.87	75.71	619.09	77.73	56.24	0.00
1375	5777	584.70	76.49	782.22	89.51	178.97	0.00
41	3696	563.64	101.72	736.00	97.48	43.34	0.00
8744	6662	299.53	88.95	597.15	85.55	65.14	0.00
309	2499	345.52	96.66	560.48	117.39	17.59	0.00
309	2452	261.75	74.39	473.84	93.93	108.83	0.00
309	2500	284.24	78.91	507.46	98.45	22.18	0.00
2273	5338	247.54	100.63	461.67	62.86	5.21	0.01
1703	6592	184.90	71.34	414.29	64.63	7.13	0.00

From this table, it is evident that for all the top 20 players, the p-value for 2-sample t-test is very small which means we reject the null hypothesis. We can infer that the average score of users selecting top players is not equal to the average score of users not selecting top players.

Key Inference out of this test is - The average scores of users who choose top players and who do not choose the top players are statistically significantly different. This implies that the users have to identify the high performers using their skill in order to win or score high.

Sixth Hypothesis- Choosing of right captain and vice-captain is very key to getting high scores in this game since the scores of captain and vice-captain are multiplied two times and one and a half times respectively. If we can test that high scorers have chosen captain and vice-captain based on the past performance of the players, then we can claim that skill is a dominant factor in forming a team mainly the captain and vice-captain.

This hypothesis is to find if there is a difference between average scores of users who are selecting top 4 players as captain or vice-captain and others who are not. There will be three samples of average scores of users who selected three top players as captain or vice-captain and one for users who did not choose any of those three players.
This hypothesis test uses the Analysis of Variance (ANOVA) to find the average scores of users selecting top players as captain/vice-captain and users not selecting them as captain/vice-captain.

Hypothesis

Ho: No difference in average scores of users
Ha: Not all the user scores are same

To carry out ANOVA, we will create 3 samples: 1. Scores of users who selected this top player as captain, 2. Scores of users who selected top player as vice-captain, and 3. Scores of users who did not select the top player either as captain or vice-captain.

We have identified top 4 players who may be chosen as captain and vice-captain. The following process was used to perform ANOVA.

- For each of these players
    - All the rounds in which these players played are selected.
    - All the users who played in the selected round are chosen.
    - For all those users based on the factor if one of the top 4 players is chosen as captain or vice-captain or not, three groups are created.
        - Group 1: Scores of users who selected the player as captain
        - Group 2: Scores of users who selected the player as vice-captain
        - Group 3: Scores of users who did not select the player as captain or vice-captain

ANOVA was performed between these three groups for all the 4 players separately. The results are provided below:

Result

1. Player: 635
   a. F-statistic = 712301.1032301941
   b. P-value = 1.1102230246251565e-16
2. Player: 119
   a. F- statistic = 165083.02614741513
   b. P-value = 1.1102230246251565e-16
3. Player: 342
   a. F- statistic = 217211.11978071192
   b. P-value = 1.1102230246251565e-16
4. Player: 6
   a. F- statistic = 21118.936498018324
   b. P-value = 1.1102230246251565e-16

ANOVA results shows that p-value is very small for all 4 instances; so, we need to reject the null hypothesis that the average scores are not equal when top 4 players are either selected or not selected as captain/vice-captain.

A couple of other hypothesis tests were performed on selection of players as captain by high and low performing users. The results, elucidated below, evidence that there is statistically significant difference between the high performers and low performers when it comes to selection of captain.

Hypothesis: Captaincy selection and performance are dependent
The following contingency table is created for two sets of users: 1. High performers (scored more than 350 points) and 2. Low performers (scored less than 150 points)
The high/low performers were then checked for selection of a specific player as captain (player number 342). The corresponding contingency table is shown below:
Contingency Table

	Player 342 Selected as Captain		Total
	Yes	No	
Uses with scores more than 350	12	27801	27813
Uses with scores less than 150	275	2456	2731
	287	30257	30544

For the above contingency table, we can perform a chi-square test of independence. Chi-square statistic value is 2685 and the corresponding p-value is almost 0. That is, there is dependency between selection of captain and performance.

Hypothesis: High performers' selection of captain is different from low performers

The following Table shows proportion of times a player is selected as a captain by high and low performing users:

Player Code	Proportion of users selecting the player as captain	
	High Performers	Low Performers
5	0.002	0.009
109	0.002	0.018
309	0.40	0.0004
342	0.0004	0.1159
612	0	0.0615
635	0.3622	0.0307

From the above table, we can see that more than 76% of high performing users have chosen either player 309 or 635 as Captain, whereas only 3% of the low performing users have chosen players 309 or 635 as captain. 11.5% of low performers have chosen player id 342 as captain, whereas less than 1% of high performers chose player 342 as captain.

Key Inference out of this test is - Based on the ANOVA results, we can claim that choosing the right player as captain and vice-captain has an impact on the average scores of the users and hence again skill of choosing a captain and vice-captain is a dominant factor than by chance choosing these players.

6. Seventh Hypothesis – Learning effect is also a very important factor to prove skill as a dominant factor. Since learning from my past team selections and then improving my future selections show that analysis and intelligence goes in team selection and it is not random.

This question will enable us to understand that in a game of skill, there will be a learning effect, that is, the users will learn the skill required for selecting the best fantasy team based on their past experience. This can be checked by developing a simple regression model in which the outcome variable is the number of wins (defined as top percentile or quartile) and the dependent variable is the number of rounds played by the user.

The null hypothesis for the regression model is defined as follows:
H_0: There is no relationship between winning a round and the number of rounds played.

The following pre-processing of the data was carried out for developing the linear regression model.

1. Winners are identified as users who are in top 5, 10 or 20 percentiles.
2. Data was extracted for users who have played at least 57 rounds.

The regression outputs are shown in below tables

Figure 5

Plot between number of rounds played and number of rounds in which the player was in top 20 percentile.

A linear trend model is computed for Winn perc 20 (user is part of top 20 percentile score) given Valid Rounds. The model may be significant at $p \leq 0.05$. The regression output is shown in **table 8 below** and the corresponding plot is shown in **Figure 5**.

Table 8

Regression output in which winning is defined as top 20 percentile

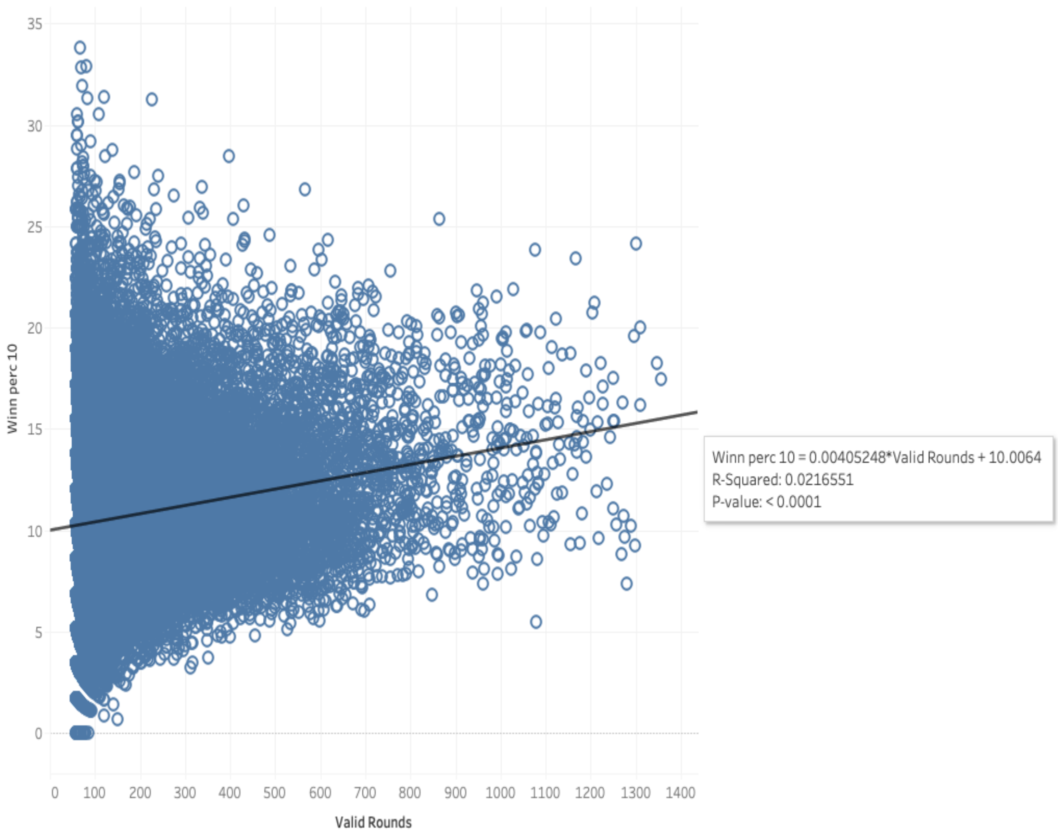
Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	55272
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	55270
SSE (sum squared error):	1.64241e+06
MSE (mean squared error):	29.7162
R-Squared:	0.0221533
Standard error:	5.45125
p-value (significance):	< 0.0001

Individual trend lines

Panes		**Line**		**Coefficients**				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 20	Valid Rounds	< 0.0001	55270	Valid Rounds	0.0059728	0.0001688	35.3857	< 0.0001
				intercept	20.2513	0.0355418	569.79	< 0.0001

Figure 6

Plot between number of rounds played and number of rounds in which the player was in top 10 percentile



Trend Lines Model

A linear trend model is computed for Winn perc 10 (user is part of top 10 percentile score) given Valid Rounds. The model may be significant at $p \leq 0.05$. The regression output is shown in **table** 9 and the corresponding plot is shown in **Figure 6**.

Table 9

Regression output in which winning is defined as top 10 percentile

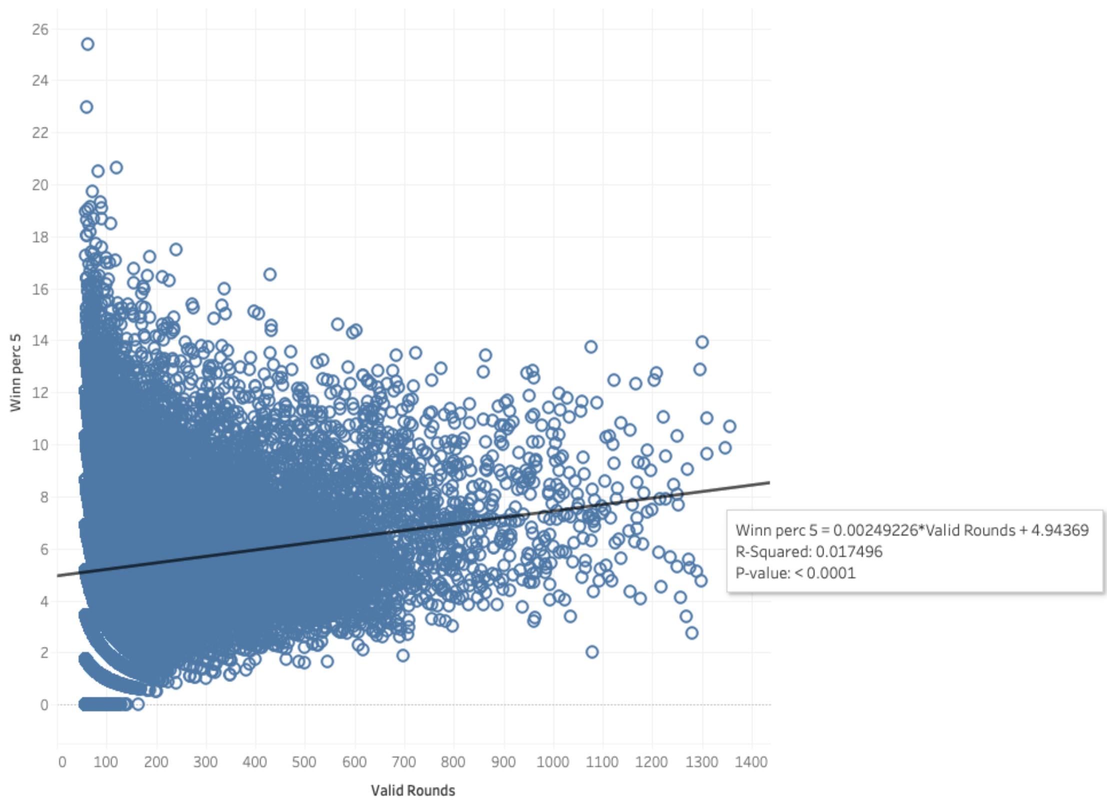
Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	55272
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	55270
SSE (sum squared error):	773864
MSE (mean squared error):	14.0015
R-Squared:	0.0216551
Standard error:	3.74186
p-value (significance):	< 0.0001

Individual trend lines

Panes		**Line**		**Coefficients**				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 10	Valid Rounds	< 0.0001	55270	Valid Rounds	0.0040525	0.0001159	34.9767	< 0.0001
				intercept	10.0064	0.0243967	410.154	< 0.0001

Figure 7

Plot between number of rounds played and number of rounds in which the player was in top 10 percentile



Trend Lines Model

A linear trend model is computed for Winn perc 5 (user is part of top 5 percentile score) given Valid Rounds. The model may be significant at $p \leq 0.05$. The regression output is shown in table 10 and the corresponding plot is shown in **Figure 7**.

Table 10

Regression output in which winning is defined as top 5 percentile

Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	55272
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	55270
SSE (sum squared error):	363808
MSE (mean squared error):	6.58239
R-Squared:	0.017496
Standard error:	2.56562
p-value (significance):	< 0.0001

Individual trend lines

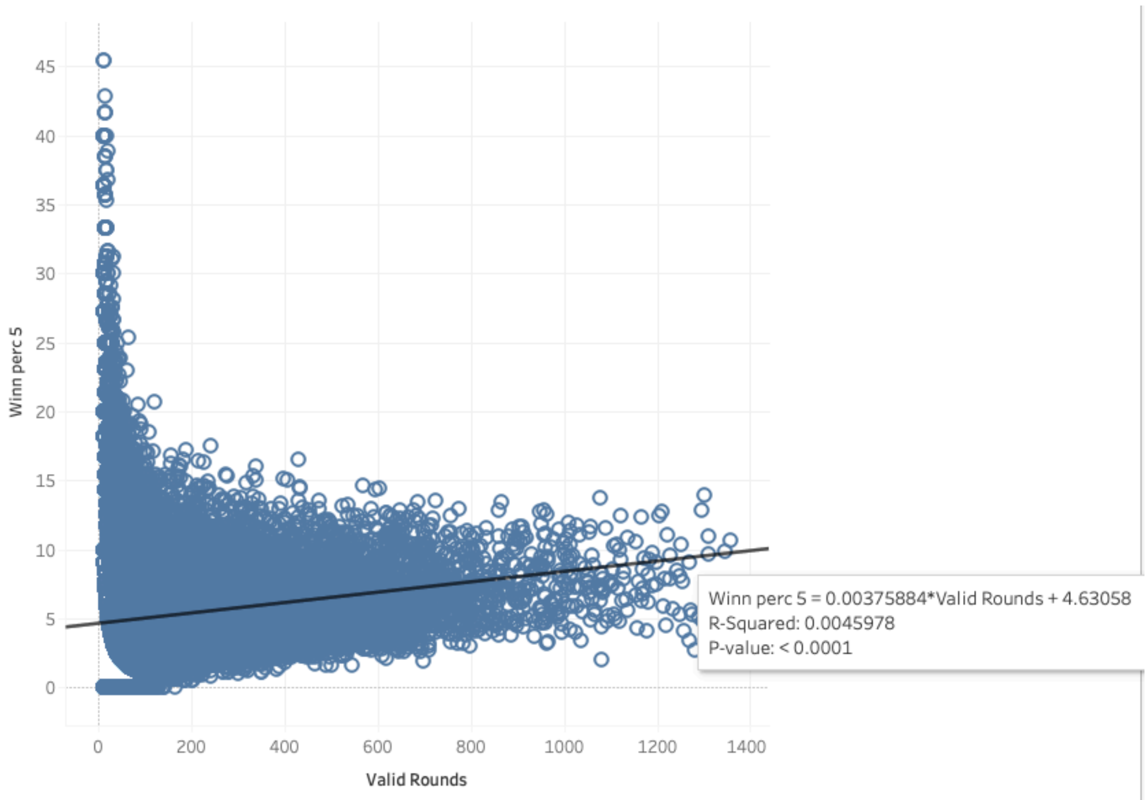
Panes		**Line**		**Coefficients**				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 5	Valid Rounds	< 0.0001	55270	Valid Rounds	0.0024923	7.944e-05	31.3724	< 0.0001
				intercept	4.94369	0.0167276	295.54	< 0.0001

In all 3 regression models, the regression coefficient for the independent variable and number of rounds (valid_rounds) is positive and statistically significant indicating that there is a learning effect.

A separate hypothesis test was also performed excluding the first 10 rounds to establish whether learning effect persists even in latter stages.

Following are the results –

- Winning 5 percentile with Valid rounds > 10



Trend Lines Model

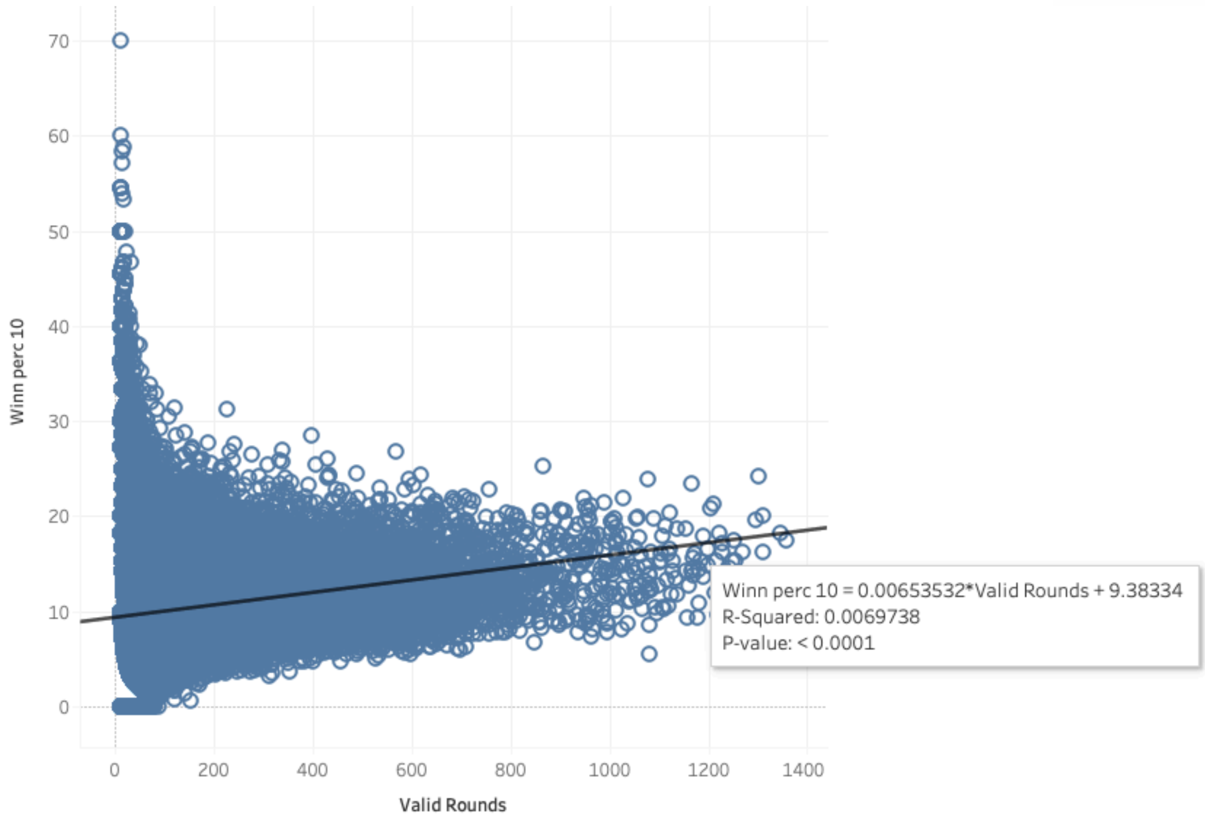
A linear trend model is computed for Winn perc 5 given Valid Rounds. The model may be significant at p <= 0.05.

Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	266660
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	266658
SSE (sum squared error):	5.80621e+06
MSE (mean squared error):	21.774
R-Squared:	0.0045978
Standard error:	4.66626
p-value (significance):	< 0.0001

Individual trend lines:

Panes		**Line**		**Coefficients**				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 5	Valid Rounds	< 0.0001	266658	Valid Rounds	0.0037588	0.0001071	35.0955	< 0.0001
				intercept	4.63058	0.0105502	438.908	< 0.0001

- Winning 10 percentile with Valid rounds > 10



Trend Lines Model

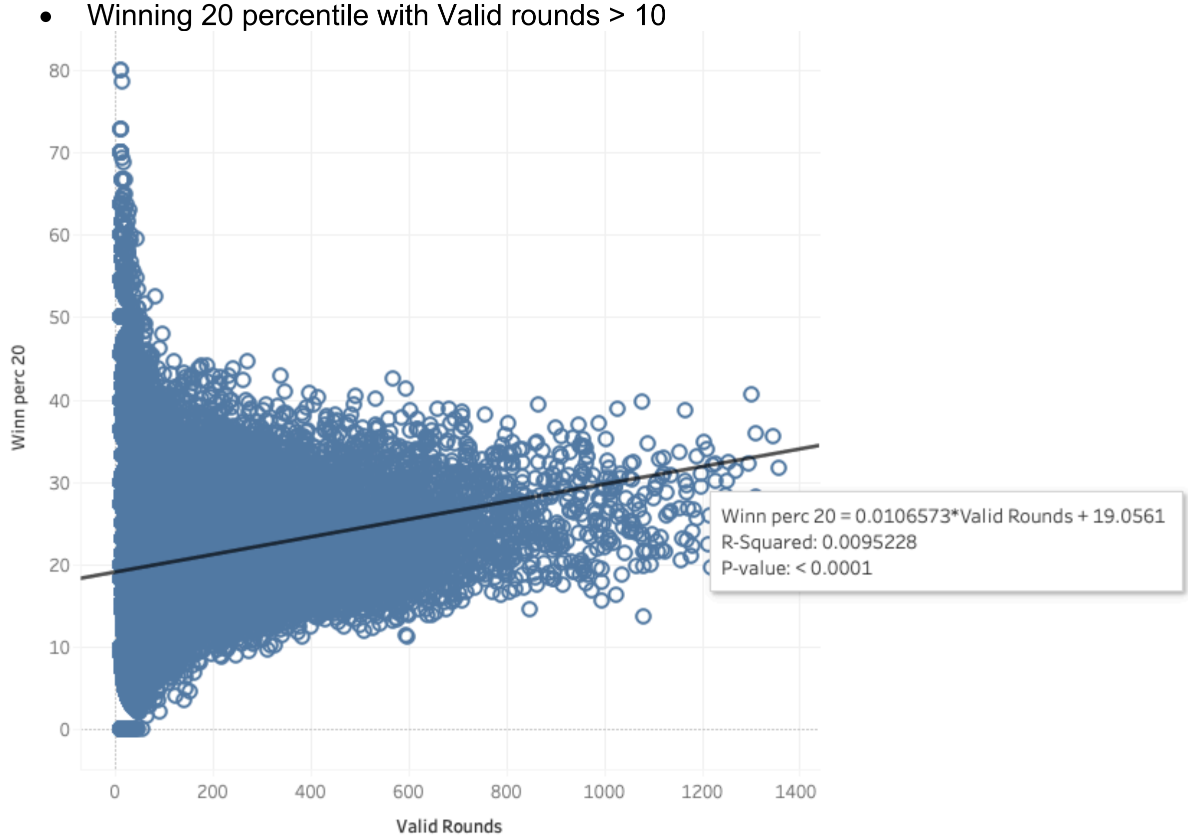
A linear trend model is computed for Winn perc 10 given Valid Rounds. The model may be significant at p <= 0.05.

Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	266660
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	266658
SSE (sum squared error):	1.1544e+07
MSE (mean squared error):	43.2915
R-Squared:	0.0069738
Standard error:	6.57963
p-value (significance):	< 0.0001

Individual trend lines:

Panes		Line		Coefficients				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 10	Valid Rounds	< 0.0001	266658	Valid Rounds	0.0065353	0.000151	43.2745	< 0.0001
				intercept	9.38334	0.0148762	630.76	< 0.0001

- Winning 20 percentile with Valid rounds > 10



Trend Lines Model

A linear trend model is computed for Winn perc 20 given Valid Rounds. The model may be significant at p <= 0.05.

Model formula:	(Valid Rounds + intercept)
Number of modeled observations:	266660
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	266658
SSE (sum squared error):	2.24237e+07
MSE (mean squared error):	84.0915
R-Squared:	0.0095228
Standard error:	9.17014
p-value (significance):	< 0.0001

Individual trend lines:

Panes		**Line**		**Coefficients**				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Winn perc 20	Valid Rounds	< 0.0001	266658	Valid Rounds	0.0106573	0.0002105	50.6333	< 0.0001
				intercept	19.0561	0.0207333	919.107	< 0.0001

The model results are significant (p value close to zero) at all wining conditions. This indicates evidence of learning effect even at latter stages.

Key Inference out of this test – This test shows that number of rounds played has a significant affect on the chances of a users to be in the top percentile of scores, hence there is a learning affect and hence skill is again a dominant factor.

SUMMARY

Based on all the tests which are done using the users, score and player data from Dream11 there are few key aspects of Fantasy Sports that we have been able to test and prove. Few of them are:

1. Users playing for money and not a free contest get higher scores proving they are playing more strategic game else.
2. In a game of chance every player has an equal chance of being selected in the team but the test done to show that randomly selected players in team lead to low scores than the scores which are obtained by Fantasy sports users prove that these users are playing more strategic game than a random one.
3. Selection of players in a fantasy sports team involves skillful assessment of a players' past performance.
4. Moreover, people who skillfully assess the past performance of players in the Captain & Vice Captain selection, tend to have higher average scores than other players.The learning effect proves that the substantial skill improvement over time which is a characteristic of a skill based game rather than a chance based game.

Based on various hypotheses tests and detailed analysis of the Dream11 data set, there is sufficient evidence to conclusively establish that Dream11 format of fantasy sports is skill dominant. Hence, we conclude that format of Online Fantasy Sports Games as offered by Dream11 requires predominant skill to play and hence a 'Game of Skill'.